# Comments on SFMWM
## by Jery R. Stedinger
**26 August 2005**

The SFWMD is to be congratulated on developing such a comprehensive and thorough water management model. The management of this system is of incredible importance regionally, as well as of great national interest. I am very pleased to see that such care has gone into the modeling of the system so that the cumulative impact of design and operating decisions can be assessed, and the impact of proposals can be predicted with some assurance. The model is a marvelous accomplishment. The documentation was in general very good. I look forward to reading the comments of other panel members and the discussion at our forth coming meeting.

I would not be doing my job if I did not develop significant comments on the model, which is really the intellectual activity of interest here. Below are lists of key and significant concerns, and also minor comments. Most of the major concerns are directed at Chapters 5 and 6. Many of the minor comments relate to the clarity of the documentation and thus are not critical. Others minor comments are of greater importance: word choice is often an editorial issue, but in other cases the choice of the words relates to the conceptual representation and explanation of the intellectual character of this activity, and thus is also worthy of careful consideration. Clearly this document is a teaching tool that needs to explain to the general public the intellectual framework and the methods the SFWMD has adopted, as well as their limitations.

## Comments of Significant Concern

**Model Purposes and Scope**
In reading the comments by the 1998 review committee comments (Loucks et al., 1998, pp. 5-6), I was struck my their concern that the objectives of the model were not spelled with sufficient clarity. In that context I consider the following comment on the stated purposes for the model to be a significant concern:

p. 3 – I do not see any difference among items #2, #3, and #4, which may be my lack of understanding of what is meant here. I would suggest that the purposes be described more clearly. I think SFWMD does #1 so as to achieve #2-4, which are basically the same. Therefore #1 is perhaps a means, and #2-4 describe the purpose. But other purposes are not described, or clearly distinguished. For example, there are certainly long-term planning issues, whereas position analysis would appear to reflect within-year operations planning efforts. As the subsequent discussion demonstrates, these activities have different character. The desire to understand the pre-development flow patterns in the region could be consider a distinct purpose.

It seems very appropriate that a clear statement of the primary purposes and intended uses of the model be made at the beginning of this document, particularly clarifying the spatial and time scales of decisions and processes that the model is intended to address. This is worth a new section between the current 1.1 and 1.2. This suggestion also appears in (Loucks et al., 1998, p. 5 bottom).

p. 4 – **Natural System Model**. it would be appropriate to say:
*Because NSM uses the same hydro-meteorological record as the SFWMM, comparison between "natural conditions" and managed systems can be made* ***MORE RELIABLY.***

One can still use the NSM for this purpose of comparing conditions without having matching meteorology. The value of using common a common hydro-meteorological record is that the comparison is much more precise. In statistical terms, one generates a paired data set, which allowed a pair statistical test. (See statistical texts cited below.)

p. 24 - Figure 2.1.1.5 – **Depressions in elevations?**
While the observations I am making are sensitive to the thresholds for different elevations, I am concerned by the depression near the center-northern boundary of the modeled region (just south of Lake Okeechobee). Does this area become a lake?

There are also blue wholes in the south eastern portion of thee modeled area several rows back from the ocean. Are these numbers correct, and if so, what is their implication?

p. 36 – **Rainfall and Climate Futures**
a) This is a very large area. Its water management could affect rainfall rates through precipitation recycling. Has this been considered? [Page 196 - bottom - discusses how reduced irrigation results in reduced ET, and could this not cause decreases precipitation elsewhere in the basin.]
b) Climate variability is now much better understood than when this project began, and climate change seems to be a certainty over the planning period for this project. => How does this project addressed these issues?

**p. 66 - Solution of PDE**
I am concerned by the described method of solving the 2$^{nd}$-order PDEs describing flow. I am not an expert on numerical solution of PDEs, though I have some experience. I am use to decreasing the time step to main stability, or reformulating a model or the numerical scheme. Clearly some time-step control is employed here.

*What troubles me is the description at the top of the page on how water movement is arbitrarily limited to maintain stability. This may yield stability, but what about accuracy?*

**While most of the report was very good and demonstrated a very high level of care. I was disappointed by several aspects of Chapters 5 and 6 addressing Sensitivity Analysis and Uncertainty Analysis.**

**Chapter 5 Sensitivity Analysis**
Perhaps we might be better off discussion this analysis rather than my writing a critical review which may be based on a misunderstanding on my part. But as a beginning for a discussion at our meeting, I venture several remarks.

It would be useful to expand the good discussion in the first paragraph of the role of the appropriate sensitivity analysis, and provide appropriate citations.

Please state clearly at the beginning what issues need to be addressed, and will be addressed, in this application by use of sensitivity analysis. What are the key questions that we hope to resolve?

p. 267 - paragraph 2 - Sensitivity analysis can be used for, and is often used for, all of the purposes listed. It is not just about parameter uncertainty. One often runs a model with different options, algorithms, data sets, and observes the differences. That is sensitivity analysis.

p. 268 – p. 278 - I found the proposed use of Singular Value Decomposition on p. 268 to be very interesting, but do not know how it was done. Note that the derivatives frequently have different units making it difficult to perform the analysis that might be suggested. I am concerned as to what was done.

p. 278, what is a parameter resolution matrix?

However did you compute what I would think of as a parameter correlation matrix (the correlation among sample estimators of the parameters) from the matrix of sensitivities?

Recall in ordinary least squares regression, solving

$$Y = X\beta + E,$$

The optimal estimator of the parameter vector $\beta$ is $b = (X^TX)^{-1}X^TY$. And the covariance matrix of the sample estimator b equals

$$\mathrm{Var}[b] = \Sigma = \sigma^2 (X^TX)^{-1}$$

where $\sigma^2$ is the variance of the components of the errors in E. The correlation between two elements i and j of b is just $\Sigma_{ij}$ /sqrt[$\Sigma_{ii}\ \Sigma_{jj}$]. So how did the SVD determine the correlations, which equal $\Sigma_{ij}$ /sqrt[$\Sigma_{ii}\ \Sigma_{jj}$] in the OLS case?

In maximum likelihood estimation the covariances among the parameters are often approximated using the second partial derivatives of the likelihood function, but this discussion did not mention second partial derivatives.

However, sometimes in nonlinear regression, the sensitivity matrix S is substituted for the X matrix in the equation above as a linear approximation of the stage function $f_i(b)$ as a function of the parameters

$$y_i - f_i(b) \sim y_i - f_i(\beta) - S\,(b - \beta)$$

yielding

$$\mathrm{Var}[b] \sim \sigma^2 (S^TS)^{-1}.$$

Is that what is being done here? One does not need SVD to employ such an approximation. If that is the issue, we might talk about the validity of this approximation of the variance, which requires that the product of the residuals times the second derivatives of f(.) is small. This is a standard method in nonlinear regression analyses. See:

Draper, N. R., and H. Smith, *Applied Regression Analysis,* 3rd Edition, ISBN: 0-471-17082-8, J. Wiley and Sons, New York, 1998.

Or, for a discussion of the linearization:

Eric W. Weisstein. "Nonlinear Least Squares Fitting." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/NonlinearLeastSquaresFitting.html

Please provide a standard and published reference describing the methods that has been adopted.

p. 268 – Section 5.2 sentence 1 – How was this done? Chapter 4 describes the calibration of the model; that does not seem to be the same as sensitivity analysis. This sentence confuses me.

p. 267 – Section 5.2 sentence 2 – Please expand upon how reasonable uncertainty range was determined for each parameter. The analysis depends critically upon these decisions and it was not clear to me what was done, and the extent to which it was consistent across parameters.

>>Later I realized that this is apparently what figures 5.2.1-5.2.2 were presented to do. But that whole discussion confused me. Please clarify what is being done and why that analysis is justified.

What criteria was employed to identify what is called a 95% confidence value? How was probability or confidence introduced?

p. 267 – Section 5.2 sentence 3 – In particular I did not see how analysis can determine "groups of parameters that are dependent on one another." What does that mean? The analysis did not vary two parameters at a time and see if the selection value for one parameters can substitute for the value of the other. (In the constraint $x + y = 5$, changes in y can be compensated for by corresponding changes in x.)

p. 267 – Section 5.2 paragraph 2 – Figure 5.2.1 and 5.2.2 – I am sure these graphs are interesting, but I do not understand what they are. Again, what is WPET? It does not appear on page 314 of the Glossary, or xvi of abbreviations. Okay, list on page 267 tells me where to look. But I could not find the term WPET on pages 44-46 which are section 2.3.1.

Could the list on page 267 be more clear as to what these parameters are, and if there is only one value, or if there are different values for each cell or some other spatial index. Yes, I could try to look each one up and resolve those questions, but I was hoping someone would do it for me and others.

Back to figures Figure 5.2.1 and 5.2.2

I think the caption of figure 5.2.1 should say:

*Sensitivity OF the average BIAS during the calibration (verification period?) TO variation of the Wetland Potential Evaporation (WPET) used in all WCAs.*

Now those variations are the different lines. But what do percentiles correspond to? Is there not one average bias (at what point(s)?) when the simulation is run over some period?

Or maybe you are computing are error for each point, and you call that a bias? So figure 5.2.1 shows me the distribution of all the errors somewhere over some period as a function of different WPET values? But what then is in figure 5.2.2 if it is not the average root mean square error somewhere over some period? I am sorry to be so confused.

The bias is described as being for the gauges in the WCAs. I assume these are stage gauges? What gauges are in the WCAs? Are they uniformily distributed spatially or should we use some weighting? (Also a concern for calibration and validation when one would also like to combine performance for different sites.) Are errors at one site averaged, and then a bias for the site computed so we have a distribution of biases over sites, or are all the sites at one time averaged, so we have a bias for each time point and a distribution of biases over time, or do we have a biases (errors) for every time point and gauge yielding a space-time distribution with the distribution shown in figure 5.2.1.

I am referred to chapter 4. Table 4.2.2.1 therre reports a bias and root mean square error for calibration and verification periods (which is used in chapter 5?), and for each stage or flow gauge. So it would appear, or I would guess, that we have a bias and root mean square error for each gauge, and the distribution is over different gauges? What was done is not clear to me.

*For the sensitivity analysis, why not just sum the mean square errors over all gauges, and have for each region a single mean square error? Why spend any time with bias: it is included in the mean square error?*

*Why are we not exploring the impact of such variations in WPET on stages in ENP?*

What is meant by the statement: "To keep both bias and rmse error small, it is preferable to have a small change in WPET." And on page 269, "To keep the modeling output valid, the recommend change of WPET is +/-10%." Why have any change in WPET? When I reached table 5.2.1, I suspect that I have not been reading about a sensitivity analysis, but rather a method for determining uncertainty ranges for the parameters! I did not know that, if it is true. Given the even values in table 5.2.1, the analysis is very crude.

p. 272 Does BCNP represents an average stage? At how many gauges?

p. 273 Figure proposes to provide the change in stage (feet) for a 100% change in parameter value? So if WPET has a calibrated value of 10, we are considering zero and 20?

Or did you actually change WPET from its calibrated value to its upper uncertainty value, and its lower uncertainty value? That is what I would have expected. See also equation 5.2.1. But we have equation 5.2.1, which does not seem to correspond to what figure 5.2.4 says is presented. I am really confused.

Figures 5.2.4 through 5.2.8 appear to be providing a clever and compact presentation of the sort of analysis that one would like. What I would like to see is the uncertainty range [y(+) – y(-)] for each stage averaged over the specified gauges, from a change in each parameter from its lower (-) to its upper (+) uncertainty range. Is that close to what is reported here?

**The sensitivity analysis is clearly trying to do the right thing, and figures 5.2.4 through 5.2.8 appear to be providing a clever and compact presentation of the results.** But what was done is not clear. Because I can familiar with nonlinear regression methods, I can imagine how the parameter correlations were obtained; but such a vision was not obtained by reading this documentation.

**Chapter 6 Uncertainty Analysis**

pp. 279 & 280 (also pp. 267 and xiv)- I think the term *confidence limit or confidence interval* is incorrectly used here in the context of uncertainty analysis. In the statistical literature, the term confidence limit has a very specific meaning (in repeated sampling 95% of 95% confidence intervals based upon the sample average with n observations will correctly contain the true mean μ), which is different from the appropriate meaning here. Thus I would suggest when the SFWMD rewrite the text using a different term, such as a probability interval, or uncertainty bands, which the text suggests.

p. 279 – the issues is not the *limitations* of model output; rather uncertainty analysis addresses the *precision* and reliability of model predictions. The model may output lots of numbers. The reliability of predictions are the issue of concern.

p. 279 – I do not have Trimble (1995a) and thus am perhaps unable to appreciate the analysis here. Do the ideas in Trimble appear in standard texts on this subject to which the documentation could refer?

Eqn 6.1.1.1. This is a standard approach that is very appropriate here. For a discussion of this method see:

Morgan, M. G. and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York, 1990.

Kottegoda, M. and R. Rosso, *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York, 1997.

Benjamin, J. R. and C. A. Cornell, *Probability, Statistics and Decisions for Civil Engineers*, Mc-Graw-Hill, NY, 1970.

However, the analysis assumes that errors in the different parameters are uncorrelated. I have not seen that demonstrated. The top of page 280 indicates that Table 5.2.3 has a correlation matrix which indicates the parameter errors are fairly uncorrelated. Currently I do not understand what was done pertaining to Table 5.2.3. The assertion may be correct.

The choice of a half-width as well as the term seemed awkward. The term in the table *half-width of 90% uncertainty band* is certainly a mouthful and does not give the sense of what analysis is about.

*Why not just report the estimated standard deviation?*

*Or else define your "uncertainty measure" to be 1.645 times the estimated standard deviation? [which is what your* half-width of 90% uncertainty band *really is.]*

Please do not use interchangeably the terms uncertainty band, and certainty band, as text indicate will be done.

p. 281 **Results discussing half-widths.**
The analysis has been done and is of great value. Can we now have some overall conceptual statement about whether these values are too large, quite small, or sufficiently small that the intended analysis with the model will be credible.

Now is the time to reflect on whether the model can meet the objectives. Loucks et al. (1998, pp. 10) make a similar request: "a discussion of the limitations of model output analysis based on the uncertainty analysis results would be useful. Given the District's experience …." They go on to suggest (Loucks et al., 1998, pp. 10), "documentation should anticipate how the model will be, and should be used…. It is better to state clearly just what the model can and can not do well…" The documentation lacks a conclusions section that would provide such guidance.

Loucks, D.P., W. D. Graham, C.D. Heatwole, J.W. Labadie, R.C. Peralta, A Review of the Documentation for the South Florida Water Management Model, Submitted to SFWMD, March 27, 1998.

**Regression Analysis**

To be blunt, it appears to me that this analysis is not needed and its execution is statistically flawed.

First, why do we need to resort to a regression analysis? I believe the standard errors from the validation effort provide a direct description of the predictive precision of the model on a data set not used for calibration. Is that not what we want?

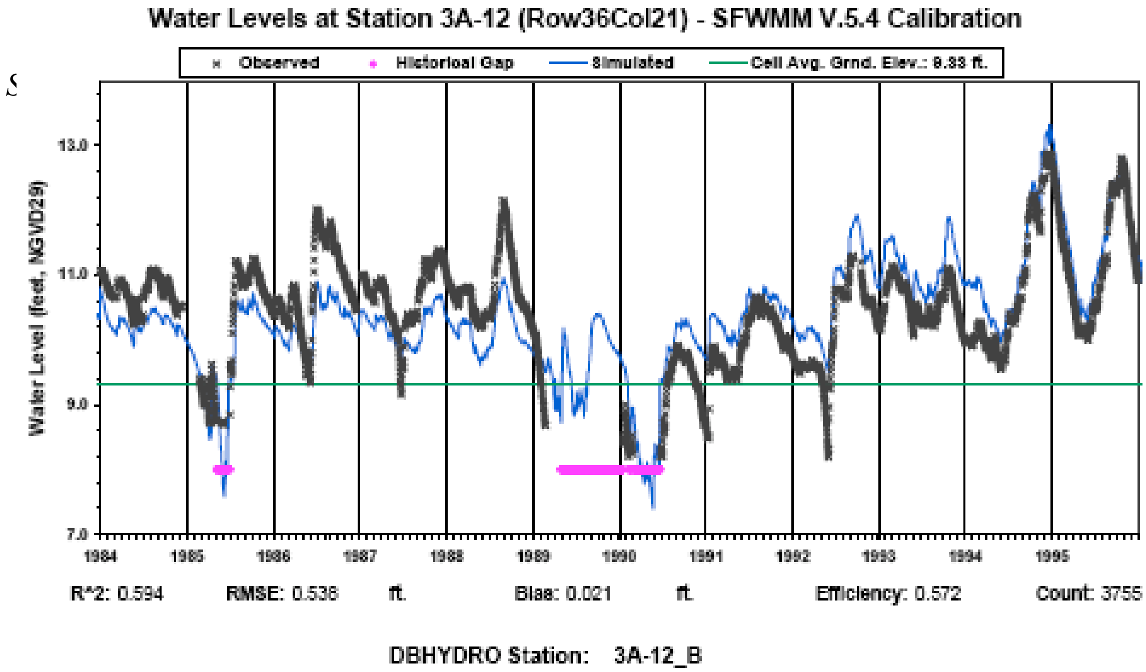
There is another concern here. Both regression and the validation statistics generally describe how well the model predicts the reported or observed values of stage and flow. If those values have errors, then the question remains as to how well the model predicts the real stages and flows. This is a difficult problem to solve. But if we have some idea of the precision of the point measurements as representations of cell averages, the issue can be addressed.

**Regression Analysis**

pp. 280-281

This analysis has serious problems. Equation 6.1.2.1 is mathematically correct, but requires several important assumptions, including that the errors are normally distributed and independent, and the underlying model is really linear. This is important and should be made clear immediately after the equation.

The assumption of normality is not addressed, but is probably not critical because we have a large sample. Independence of the residuals is not addressed, and appears to be unlikely is many cases, as is shown by figures in appendix C, such the included here. Clearly there are long series of positive and negative residuals as over a period of many months the model continues to over-predict, and at other times to under-predict, the value of the observed flows. I assume this occurs because the rainfall gauge network has failed to capure the actually volume of rainfall, and it takes awhile for the accumulated difference to work its way out of the system. This kind of correlation in the errors is to be expected.




Figures 6.2.1 demonstrates that the linear assumption is invalid. The model clearly fails at the low end.

Thus while use of a regression model may appear appealing for quantifying the precision with which the simulation model can predict observed flows (or have the authors done it the other way around; definition of Y is unclear and figures would suggest that the authors are trying to predict the simulation values with the observed values, which would be strange), which requires several assumption that do not seem to have been checked, and probably are invalid. But there seems to be no need for the analysis in the first place because that the calibration analyses provided the equivalent measure of precision.

## Minor Comments

p. xiii – what is a "nicely" configured workstation?

p. xiv – Rather than saying the model calibration is used to *reinforce*, would it be more appropriate to say the model is used to *evaluate* …

p. xiv – I think the term *confidence limits* is incorrectly used here in the context of uncertainty analysis. In the statistical literature, the term confidence limit has a very specific meaning, which is different from the appropriate meeting here. Thus I would suggest when the SFWMD can, that the text be rewritten using a different term, such a probability interval.

p. xiv – glossary is WONDERFUL.

p. 1 – last paragraph – text would be clearer if the location of the EAA and five WCAs was described.

p. 4 – line 7 from thee bottom – *maintaining* is the wrong word; maybe respecting?

p. 7 – When you say *dominance* of evapotransportation, overland flow and groundwater movement, this is in constract to what? That is, what is left that it dominants?

p. 7 – I found the sentence confusing that says: *The model is conceptualized … for three different major geographic areas (1) for … (2)….. and (3). The model is conceptualized … for three different major geographic areas (1) for … (2)….. and (3).* Numbers would make the text clearer.

p. 10 – Rate of evaporation from below the land surface generally depends upon the type of plant and their root system. Some discussion of plants and their root system would be useful? Are their periods when they are dormant?

p. 19-20 Last bullet and first bullet – Why are some cells excluded? Do they not need elevations assigned?

p. 37 – If rainfall is highly variably spatially (thunderstorms in summer months), than applying one station over a large area gives to big a variance to the total rainfall depth is sparsely gauged areas.

p. 38 – **Why were some very large values recorded?** Are these really errors, or isolated and extreme point measurements? If they are real and we exclude them, do we bias our estimate of the mean precipitation?

p. 42 – **constructive suggestion – Figure 2.2.2.3 is very nice.** The mean is easy to understand, but the meaning of +/- one standard deviation is not. I suggest that you plt 10-perecentilie, mean and 90-perecentile, or some other percentiles. This is would be more informative that +/- one standard deviation: the date is clearly not normal in Nov-May, and percentiles would help us see what the real distribution looks like.

p. 64 – equation 2.4.1.4 – typo?

p. 64 – it seems to be that rooted vegetation and islands are going to give a Mannings n value different that generally experienced in open-channel flow modeling. If so, such a warning would be appropriate.

p. 70 – figure 2.4.2.2 – Gives n as a function of *ponding depth*, not *depth*.

p. 72 – please define the term *borrow canal.*

p. 96 - Figure 2.7.2.3 – Can you tell us why some points are SO bad? Some of these points are really bad, perhaps invalidating ordinary least squares as an appropriate statistical technique, and I mean that.

p. 97 – figure 2.7.2.4 – Not as bad as figure 2.7.2.4, but still terrible. Same comments apply.

p. 98 – Figure 2.7.2.5 – This may be worse than figure 2.7.2.3; what is going wrong here? How should this process be modeled? Should we maintain the variability that has been lost using

Maintenance Of Variance Methods (MOVE)? The straight-line fails to represent the character of the data. Or is this a small and relatively unimportant process.

p. 96-98 – Figures – It would be more standard, and thus much easier for me to understand the figures, if predicted were on the horizontal azes, and observed on the vertical access.

p. 101 – Why no GW flow at Gulf of Mexico at southwestern boundary? Seems unreasonable that there would be no groundwater flow to the Gulf.

p. 105 – with two tridal cycles per daayt, and a 1-day model time step, how are these processes reconciled?

p. 108 – NOT ALL aspects of current and future proposed operations are included. Many are not important. Documentation needs a better prespective.

p. 119 – first paragraph – significance options for the use of the model are hidden here, and should be in a more obvious location.

p. 119-120 – Figures 3.1.4.1 and 3.1.4.2 is unreadable. What do these FORTRAN names mean?

p. 178 – Items 1-2-3 confuses me, even with examples. How is offset different from translation? Because of order?

p. 185 – I found the listing of three consumption categories to be inconsistent and confusing. How is irrigation a source of water? Where does the irrigation water come from?

p. 202 – line 4 – instead of *detailed* say *described.*

p. 208 – paragraph 2 section3.6.2 – what does it mean that water must be replaced?

p. 212-213 - I have heard the term position analysis applied to this sort of analysis before. The method warrants a clearer presentation that found at the bottom of page 213.
Pleased provide one or more citations to the literature describing the use of the idea by others. Maybe those below?

> Hirsch, R. M., Synthetic Hydrology and Water Supply Reliability, *Water Resources Research*, 15(6), 1603-1615, 1979.
>
> Day, G. N., Extended Streamflow Forecasting Using NWSRFS, *Journal of Water Resource Planning and Management*, vol. 111, no. 2, pp. 157-170, April 1985.

p. 215 - Figure 3.6.2.6 - I got lost and was unable to follow the derivation of this figure. Why do the sequences no longer start at the top of the supply-side management zone?

p. 223 Figures in my option have been drawn with the axes reversed. Data should go on the vertical axis, and model predictions on the horizontal axis.

**Simulated values in both figures seem much to large in both figures. Can you explain?**

pp. 223-230 – Please provide calibration and verification goodness-of-fit statistics. These are eventually defined on page 235, but are needed earlier in this section.

p. 267 Say
*Sensitivity analysis is to be distinguished from uncertainty analysis.*

p. 267 - I do not like the use of *confidence* as a term describing uncertainty as used here.

p. 236 – Can you pleased provided a clearer explanation of the difference between Eff and R^2 than is provided by 4.2.1.5. Please.

pp. 253 & 255 – figures 4.3.1.2; 4.3.1.5 would be improved by use of quantiles rather and +/-; see earlier comment.

p. 279 - Eqn 6.1.1.1 has a P-subscript in the wrong place.